

Classification of database applications

- Two kinds of processing :
 - Transactional -> OnLine Transactional Processing (OLTP)
 - Ability to collect and manage data
 - Analytical -> OnLine Analytical Processing (OLAP)
 - Ability to create information from data
 - Analyze operational data to create reports and support decision making
- Online means end-users expect fast answers to their queries

OLTP vs OLAP

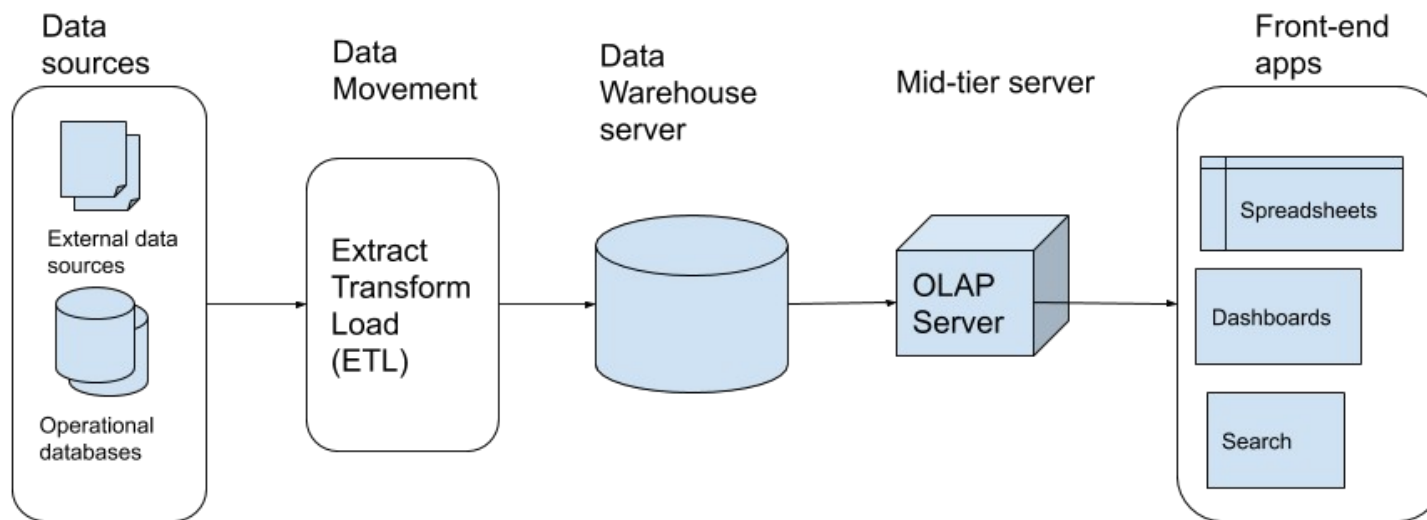
	OLTP	OLAP
Read pattern	Few tuples fetched by a key (index)	Large number of records are accessed
Write pattern	High frequency of small transactions	Low frequency of very large transactions, Bulk import, Data streams
Latency	Low	High
Dataset sizes	GB to few TB	TB to several PB
End-users	Many non specialized	Few specialized (Business analysts, Managers)

Kinds of analytics

- Descriptive analytics: what happened?
 - *Did we run out of beer cans in store S last month?*
- Diagnostic analytics: why did it happen?
 - *Get contextual information that last month was the beginning of the soccer world cup*
- Predictive analytics: what will happen?
 - *Use external knowledge of coming sport events to know when to stock up beer cans in some stores*
- Prescriptive analytics: how can we make it happen?
 - *Sponsor sport events to sell more beer cans*

Data storage

- “A Data Warehouse stores and manages data. OLAP transforms Data Warehouse data into strategic information” (OLAP Council)



Analytical processing

- Performed by OLAP and/or data mining components
- There are several kinds of OLAP systems
 - Multidimensional OLAP (MOLAP)
 - **Relational OLAP (ROLAP)**
 - Hybrid OLAP (HOLAP)
 - And the emerging Graph OLAP (GOLAP)

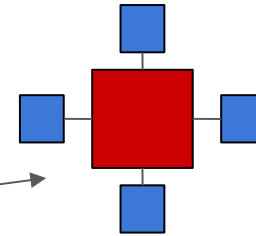
ROLAP - Schema

- Distinction between **fact** (e.g., item sales) and **dimension** (e.g., store, customer) tables

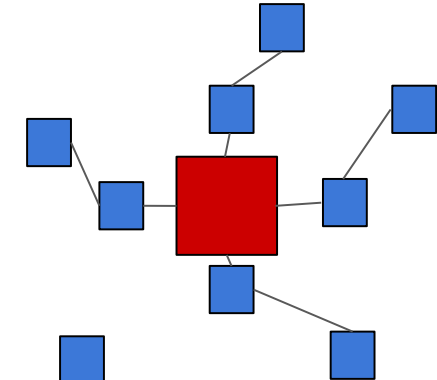
ROLAP - Schema

- Distinction between **fact** (e.g., item sales) and **dimension** (e.g., store, customer) tables
- Schemata

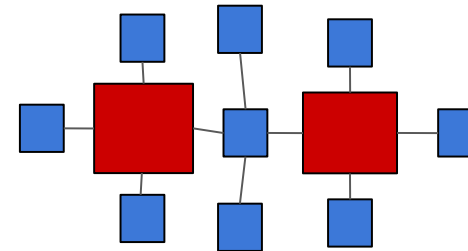
Star schema



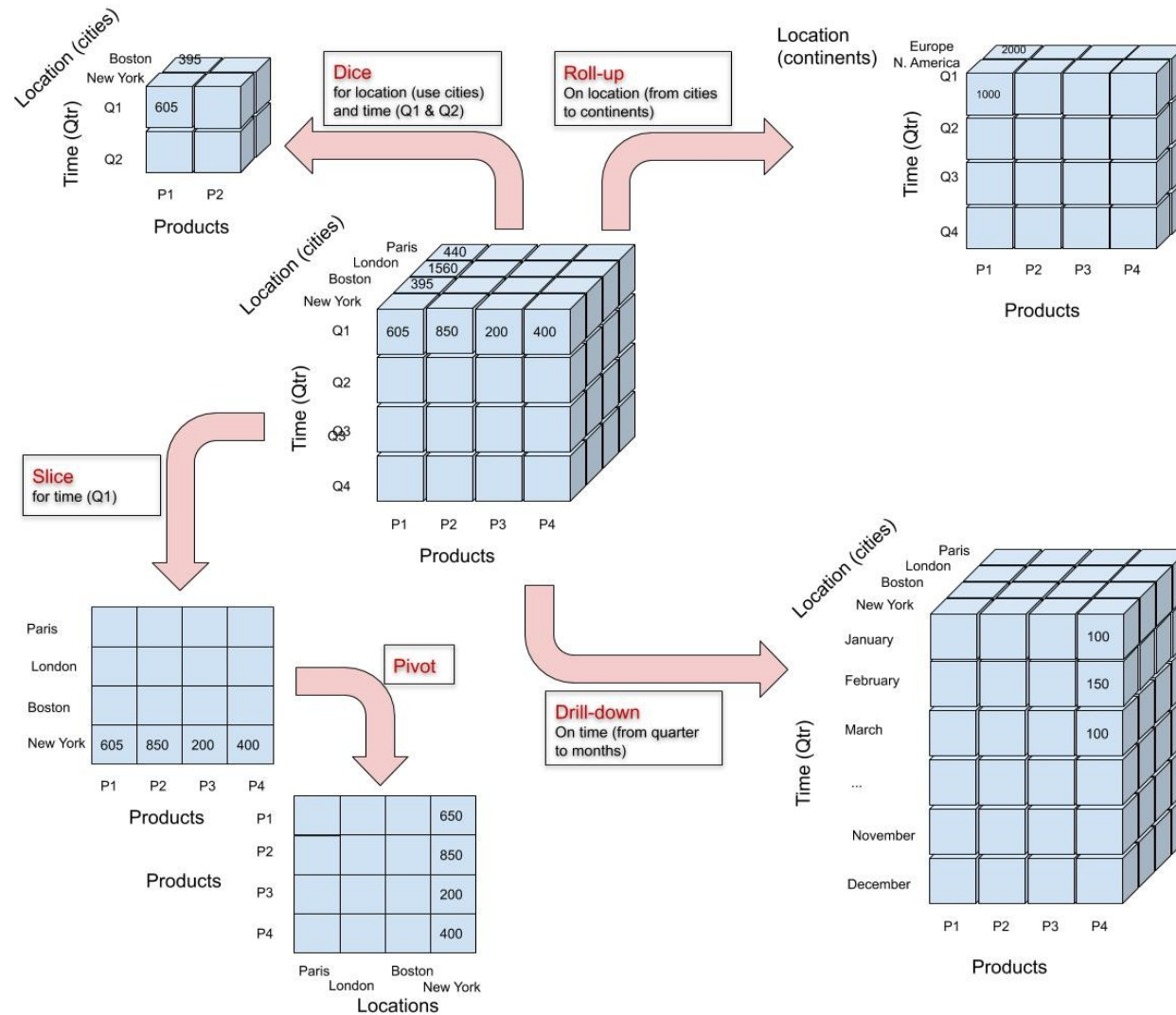
Snowflake schema (multiple levels of dimension tables)



Constellation schema (multiple fact tables)



ROLAP - SQL extensions



OLAP - Views

	Virtual views	Materialized views
DB system	OLTP	Data Warehouse
Structure	Logical table	Logical table
Persisted in DBMS	No	Yes
Latency	High	Low
Cost of updates	Low	High

OLAP Storage

- Column-oriented storage
 - Stores tuples column-wise (not row-wise)
 - More efficient for select queries retrieving a subset of the tuples of some tuples
 - Limited on writing operations (have to access several files to insert/delete a tuple)

Row Store v. Column Store

Record #	Name	Address	City	State
0003623	ABC	125 N Way	Cityville	PA
0003626	Newburg	1300 Forest Dr.	Troy	VT
0003647	Flotsam	5 Industrial Pkwy	Springfield	MT
0003705	Jolly	529 S 5th St.	Anywhere	NY

Record #	Name	Address	City	State
0003623	ABC	125 N Way	Cityville	PA
0003626	Newburg	1300 Forest Dr.	Troy	VT
0003647	Flotsam	Industrial Pkwy	Springfield	MT
0003705	Jolly	529 S 5th St.	Anywhere	NY

OLAP - Data storage

- Compression

- More possibilities to compress data on columnar storage than row storage
- Solutions

- Run-length encoding
- Delta encoding
- Bit encoding

